

**Best
Available
Copy**

ADA 285362

VOICE CHANNEL AUTHENTICATION

A Proposal for Scientific Study and a Special Vocoder System Channel

LAWRENCE V. KRIGER

PROJECT 4610

FEBRUARY 1960

COMMUNICATION SCIENCES LABORATORY
ELECTRONICS RESEARCH DIRECTORATE
AIR FORCE CAMBRIDGE RESEARCH CENTER
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
BEDFORD MASSACHUSETTS

ABSTRACT

Military requirements for voice channel authentication are cited. State-of-the-art assessment of speech recognition and vocoder techniques is related to voice authentication problems and requirements. Two in-house devices are suggested for feasibility and system approach studies. Non-military useful benefits are also briefly discussed.

iii

Accession For	
NRIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

VOICE CHANNEL AUTHENTICATION

A Proposal for Scientific Study and a Special Vocoder System Channel

1. INTRODUCTION

Military requirements for authentication of voices transmitted over special communication channels are of a limited but extremely important nature. The requirement for authentication when verbal orders for massive retaliatory destruction must be transmitted is easily appreciated as one of several military needs. The paper principally concerns itself with the problem of exactly identifying a person who may be using such a channel and techniques will be described for dynamically accomplishing such a "pseudo-fingerprint" mission. Related problems and applications are also discussed. The ability to cope with the general problem of authentication will be based on general state-of-art techniques plus some novel considerations.

In particular, two in-house programs are suggested for feasibility and systems development studies. The first program discusses a study and development of speech characterizing patterns which could be performed by the vocoder-computer-pattern-storage techniques proposed by C.P.Smith⁽¹⁾. The information on voice pattern statistics and coding, which is available or programmable from Smith's equipment, is shown to be implementable for authentication studies and instrumentation. The second program suggests the utilization of the Polar Coordinate Converter in the manner shown by F. Vilbig⁽²⁾ for studying the phase contribution and invariance of such phase for speech polar diagram patterns.

Such studies would be useful for ascertaining the feasibility of using speech spectrum phase information for distinguishing a particular voice from a random set. The success of such studies would be an important clue as to the potentiality of voice authentication techniques, in general.

The similarity of difficulties associated with exactly authenticating a voice which has been transmitted over a communication channel (authentication) and of establishing the identity of a stored voice message, such as might appear on a recorded tape (identification), are considered from the viewpoint of the intrinsic speech analysis which must be performed. The identification aspect is thereafter assumed part of the authentication problem

2. PROBLEMS OF VOICE AUTHENTICATION

A basic problem of authentication is the degree of reliability that may be attained. The question may well be posed "Is it possible to identify a voice detected from a radio-communication channel with the same reliability realized in fingerprint identification." A more practical aspect of this problem would be the question of how much reliability must or should an authentication channel have. The answer to these questions are not known, but the state-of-art in speech analysis, especially with respect to pattern recognition and pattern coding, has led the author to believe that the time is now ripe to evaluate these

problems. He is further led to believe that at least a good degree of reliability may be achieved for voice authentication. This feeling is partially confirmed by the limited results achieved by H. Dudley and S. Balashek in their Automatic Speech Recognizer experiments⁽³⁾. Here their machine responded properly only when it was verbally addressed by a particular voice (the designer's).

The possible problems involved in consistently identifying a voice over a radio link will be varied. Biological changes in the talker's vocal passages, such as hoarseness, extracted teeth etc. could play havoc with the system. Instrumentation environment and transmission distortions must also be taken into account. These problems may be taken care of in the instrumentation and transmission techniques of the special authentication channel. A problem which is more basic to the overall situation and which cannot be handled by instrumentation techniques is that of inherent voice pattern consistency. It must first be shown that vocal "fingerprint" characteristics do exist and are maintained under varying conditions. It is thus concluded that a program of scientific studies and special system instrumentation must be initiated to gain a better insight to these problems. The need for a broader approach to speech recognition (which is germane to authentication) was voiced by E.E.David⁽⁴⁾ at the recent AFCRC seminar on speech compression. This is indeed true in the specialized area of authentication since prior studies in speech recognition have been

oriented toward common (invariant) attributes of voice signals rather than particular (personal) attributes. Studies of optimum criteria for authentication are seriously lacking. Such studies could point to identification characteristics other than those present in a speech spectrum pattern. Therefore a concerted effort, including those related scientific disciplines which might profitably contribute, is necessary to evaluate the overall problem, eliminate the many "garbage" techniques and assure a near future solution.

3. VOCODER APPROACH

The state-of-art of vocoders was well described at the recent AFCRC seminar on speech compression (session B of the seminar proceedings). Performance data would indicate that digitized vocoders could be profitably used for secure transmission of information. Here the voice information would be digitally coded, and the interposition of a special coding black box could give privacy to a particular channel. A serious disadvantage of using such a vocoder over a sensitive military command channel (requiring authentication) is that the synthesized voice at the receiver terminal could never recapture the exact original pitch of the sender, ⁽⁵⁾ allowing limited but not authentic recognition of the received synthesized voice. This situation could allow transmission of vital verbal commands which might not be "authentic". Authentication procedures must prevent such situations from arising.

It might be profitable at this point to reflect on the difference between listener voice recognition and receiver voice authentication. Recognition here implies that the voice "sounds like" a particular person, where apriori information of the sender's voice is required by the person receiving the message. Authentication requires that the sender's voice is positively identified and that this type of identification may or may not require the person receiving the message to be familiar with the sender's voice. The authentication technique could be to store pattern information regarding the particular sender's voice at the receiving terminal. Here it would compare the incoming digitized voice patterns with prestored information in order to make an authentication decision. In the case of marginal authentication decisions, repeat transmissions or special questions might be requested from the sender to improve the correlation of authenticity.

The vocoder system properties which best lend themselves to authentication purposes are the pitch channel parameters and the time-multiplexed amplitude-quantized spectrum patterns which quasi-continuously characterize the speech spectral energy. Computer programming procedures would best be suited for analyzing such parameters or pattern streams with respect to a particular talker, and information comparison and storage techniques utilized to assign an authentication weighting to a channel user's voice patterns. Techniques for deemphasis of invariant patterns and enhancement of singular events is necessary.

An in-house device for authentication studies and instrumentation development will now be discussed. The approach by C.P. Smith⁽²⁾ of statistically encoding and decoding speech in real time for analyzing the acoustic domain of speech (using a digitized vocoder in conjunction with data processing equipment) could be modified for authentication purposes. His system, in part, can encode speech spectrum patterns from an 18 channel vocoder to gain statistical information about the invariant and transitional speech patterns which are applicable to a large population of talkers. The data is programmed into a special purpose computer, in real time, and automatically acquires a statistical description of the time-multiplexed frequency amplitude patterns of large speech samples. It should be possible to use similar techniques for authentication requirements. As against normal use of the system, the low correlation of certain phonetic patterns could now be used to enhance authentication of a particular voice. The single talker case would thus require less speech patterns than the normal use of digitally coded speech transmission. This could mean that an authentication channel might be made higher fidelity than a normal channel assuming equal transmission terminal equipment and bandwidth. This is entirely within keeping of the authentication motif, namely highly recognizable synthesized speech. It could well turn out from authentication investigations that optimum digitized speech transmission and synthesis methods, for limited cases, would be to use authentication techniques.

The improvement of digitized vocoder operation or bandwidth reduction aspects should not be tried before a more basic consideration has been studied. This is the question of voice discrimination against non-authentic channel users. It must first be shown, for instance, that a competent ventriloquist could not fool the authenticating system. To gain some special information about the ability of a voice spectrum to remain critically invariant for authentication purposes, another in-house study-approach is suggested in the next section.

4. SPEECH PHASE PATTERN STUDIES

A speech pattern is not only representative of its defining sound but also includes special characteristics of the person uttering the sound. Up to now only vocoder type patterns have been discussed, such patterns typifying the analog portrayal of the speech spectrum power density. These patterns are accordingly used to synthesize a talker's voice at a remote terminal. It would be useful, for authentication purposes, to examine other characteristics of the speech wave such as its component phase attributes. There is, in-house, a device called the Polar Coordinate Converter which allows the examination of complex spectra in polar pattern form. It turns out that when speech is examined by this device ⁽²⁾ the resulting pattern is highly sensitive to the phase analysis of complex speech sounds.

The Polar Coordinate Converter is essentially a device which portrays complex spectrums on an oscilloscope in polar form. It was designed

primarily to measure the phase and amplitude distortion which might obtain within a modulated carrier after it had been propagated through a non linear space. It provides means for photographing the spectrum patterns in polar or rectified Cartesian form. The Cartesian patterns may be photographed on 35mm film continuously or in a time-multiplexed column array. The multiplex techniques allows for compact assessment and rapid evaluation of massive amounts of pattern data such as would be necessary for a thorough speech analysis.

Since speech spectral components are essentially harmonic in nature during vowel utterances, the polar mode of the device realizes an almost infinite variety of interlooped patterns for a given set of sounds. For one given sound, a different pattern for various talker's uttering the sound may be realized. The numerous patterns amongst talkers, for a given sound, are due to the varied phase and frequency of the spectrum components. This may be easier understood by practically considering the human differences in vocal pitch, teeth positions, tongue sizes, oral cavity resonances and muscular actions which affect pitch overtones. It is not presently known how well phase patterns may characterize or discriminate a particular voice. Scientific information is very meager on this point since, classically, phase information of speech spectrum components has been neglected since it has been shown to contribute practically nothing to speech understandability (first pointed out by Helmholtz in 1877). The availability and use of the Polar Coordinate Converter should facilitate getting such phase information and

evaluate its utility for authentication purposes.

5. OTHER APPROACHES

Phase and spectrum are not necessarily the only or best parameters for speech-identifying pattern-coding purposes. Some desirable parameters are cumbersome to handle or difficult to extract. From the point of view of problem efficiency, a more sophisticated method which would be naturally suited for authentication is that which recognizes important phonemic or certain other linguistic features of a particular voice. Methods such as this ⁽⁶⁾ are still plagued by problems in phoneme classification and resynthesis for normal speech synthesis requirements. It is possible that by combining some important phonemic features in parallel with special spectrum pattern information ⁽⁷⁾ that the ideal pattern classification method for authentication may be attained. It is very likely that such a synthesis method would minimize its problem plague, and could be made very successful for synthesis purposes of the single talker case encountered in an authentication channel. Other aids for authentication would be the computer speech programming studies currently under investigation at Bell Telephone Labs. and M.I.T. (see also papers in AFCRC proceedings of symposium on speech compression and processing AFCRC TR-59-198, Vols I and II).

The requirements for authentication are not necessarily limited to military usage and for such non-military cases, depending on the degree of authentication required, other novel approaches might prove of

of greater utility. For instance, one might use patterns which included dynamic physiological or biological information of the channel user. Such pattern features might be derived from lie detector equipment responses, physiological responses or other characterizing features that could be selected more desirably by medical rather than electronic scientists. Speech researchers could look further into such abstract defining parameters as voice vibrato, breathing, explosiveness of utterances etc. Linguists, statisticians and speech therapists may well provide interesting and valuable cues to this special aspect of authentication.

6. OPERATIONAL USES

A voice authenticating channel may be utilized in several ways for military requirements. A prime requisite would be to authenticate highly sensitive command channels that might be tied into a global communications network. Here the requirement for preventing enemy encroachment into the channel for transmitting non-authentic commands is easily met. The authenticating channel would easily sense such foreign voices.

Normally the authenticating channel would be one of a multi-channel carrier system. Should the situation require, it would be possible to make the authentication channel transmit nonsense information in the case of a non-authorized user. As a countermeasure, it could be used as an additional constraint on a coded channel. This might be done in several ways. Noteworthy would be to invert the patterns in such a

manner that there would be no correlation between regular channel patterns and those of the authentication channel.

"Fingerprint" identification of voices could open up a whole new field of application of military requirements for recognition or identification. The combat problem of airplane identification (IFF) could be augmented by propeller or jet signal identifiers. This would be for situations where IFF was inoperative but radio contact or acoustic signal detection was possible. Unfamiliar voices in such air-to-ground transmissions could be recognized by comparison with centrally pre-stored information about the pilot in question. The whole idea of secure voice transmission might have to be reevaluated once "fingerprint" identification of voices became a practical matter.

New military operational requirements for authentication and identification may appear when manned space flight becomes a reality. It is difficult to predict what exigencies will prevail at such a time. For instance, an unlikely, but nevertheless possible situation might be the case of voice communication from an environment whose composite air density might be different from that on earth. This could cause the pitch of the talker to change sufficiently so that his voice might not be recognizable. By using techniques acquired for voice authentication, it should be possible to synthesize a recognizable voice from its distorted pattern set. The general problem of transmuting voices is intriguing and might find other military usefulness. This

would be in the realm of countermeasures where talker concealment rather than authentication would be involved.

Non military applications should also be considered for the useful benefits that might be derived. For instance, identification of taped voices used as evidence in courts of law could now be made certain by analyzing the taped voice with that of the person in question. Many other commercial applications will certainly arise once the basic techniques have found continued successful application.

7. SUMMARY

Vital military needs and uses for voice channel authentication have been discussed. It is indicated, from state-of-art considerations, that voice authentication is possible, but that the degree of authentication attainable must be investigated. Two in-house programs have been suggested for feasibility and system approach studies. Non military useful benefits of authentication procedures have also been discussed.

REFERENCES

1. C.P.SMITH, An Approach to Speech Bandwidth Compression, Proc. of Seminar on Speech Compression and Processing, AFCRC TR-59-198, Vol. II, Sept 1959.
2. F. VILBIG, Visible Speech-Rotary Field-Coordinate Conversion Analyser, IRE PG Trans. on Audio, AU-2 3 p76, May-June 1954.
3. H. DUDLEY and S. BALASHEK, Automatic Recognition of Phonetic Patterns in Speech, J. Acoust. Soc. Am. 30 8 p721 Aug. 1958.
4. E.E. DAVID, Some Problems in Speech Recognition, Proc. of Seminar on Speech Compression and Processing, AFCRC TR-59-198, Vol II, Sept 1959.
5. L.V. KRIGER, Pitch Extraction for Speech Synthesis with Special Techniques for Use in Digitized Bandwidth Compression Systems, AFCRC-TR-59-116, ASTIA Doc. No. AD214449, March 1959.
6. Northeastern U., Final Report on Visual Message Presentation, AF Contract 19(604)-1039 Item I, 31 March 1957.
7. Northeastern U., Final Report on Speech Analysis, AF Contract 19 (604)-3465, AFCRC TR-59-187, 31 Aug. 1959.